

ODIHR – NHRI Academy

2022: AI and Human Rights

20-24 June 2022, Tirana, Albania

Day 1 – Session 1

WHAT IS AI AND HOW DOES IT WORK? FACTS AND MYTHS

Trainer: Francesca Fanucci



European Center for
Not-for-Profit Law



www.ecnl.org

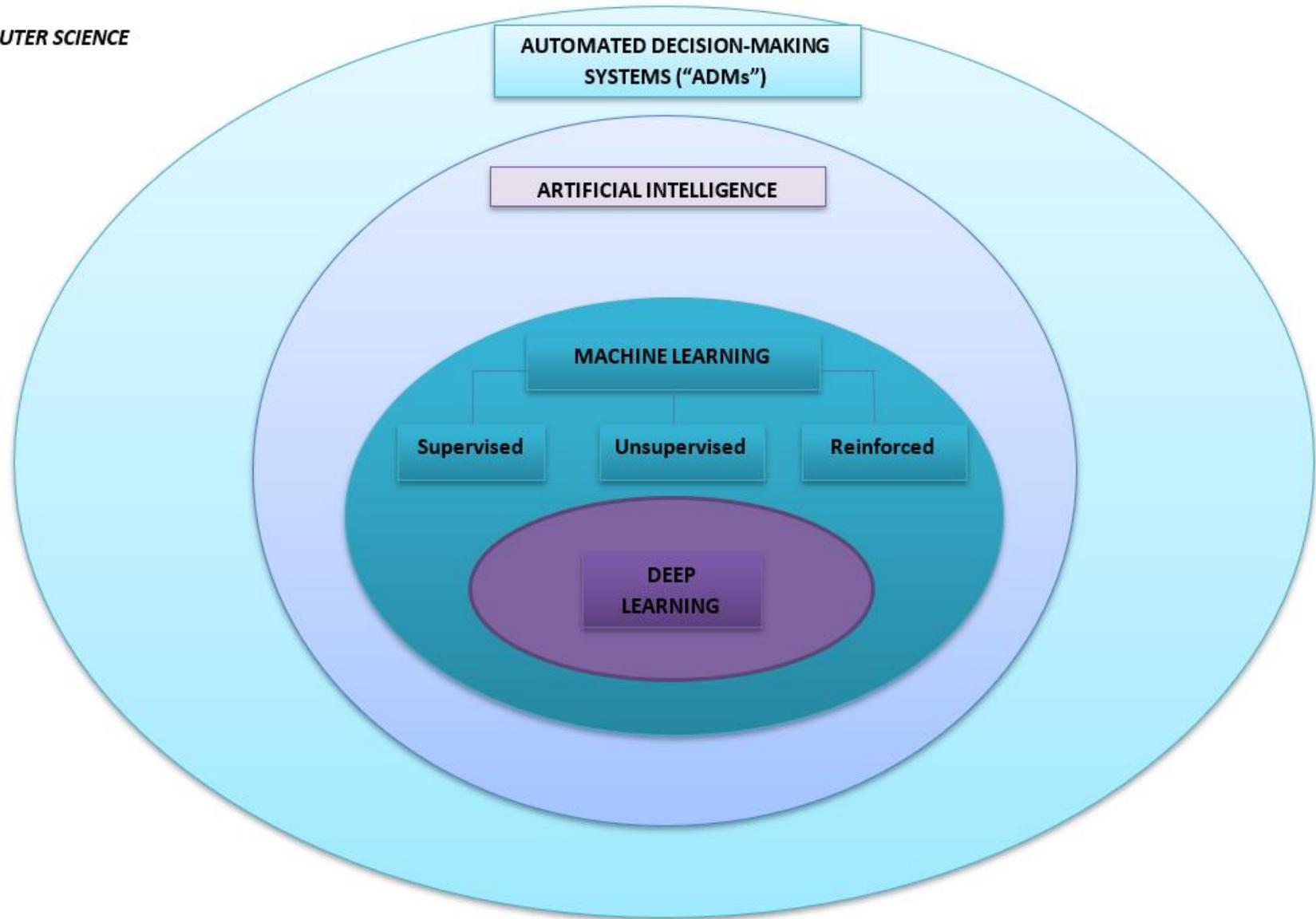


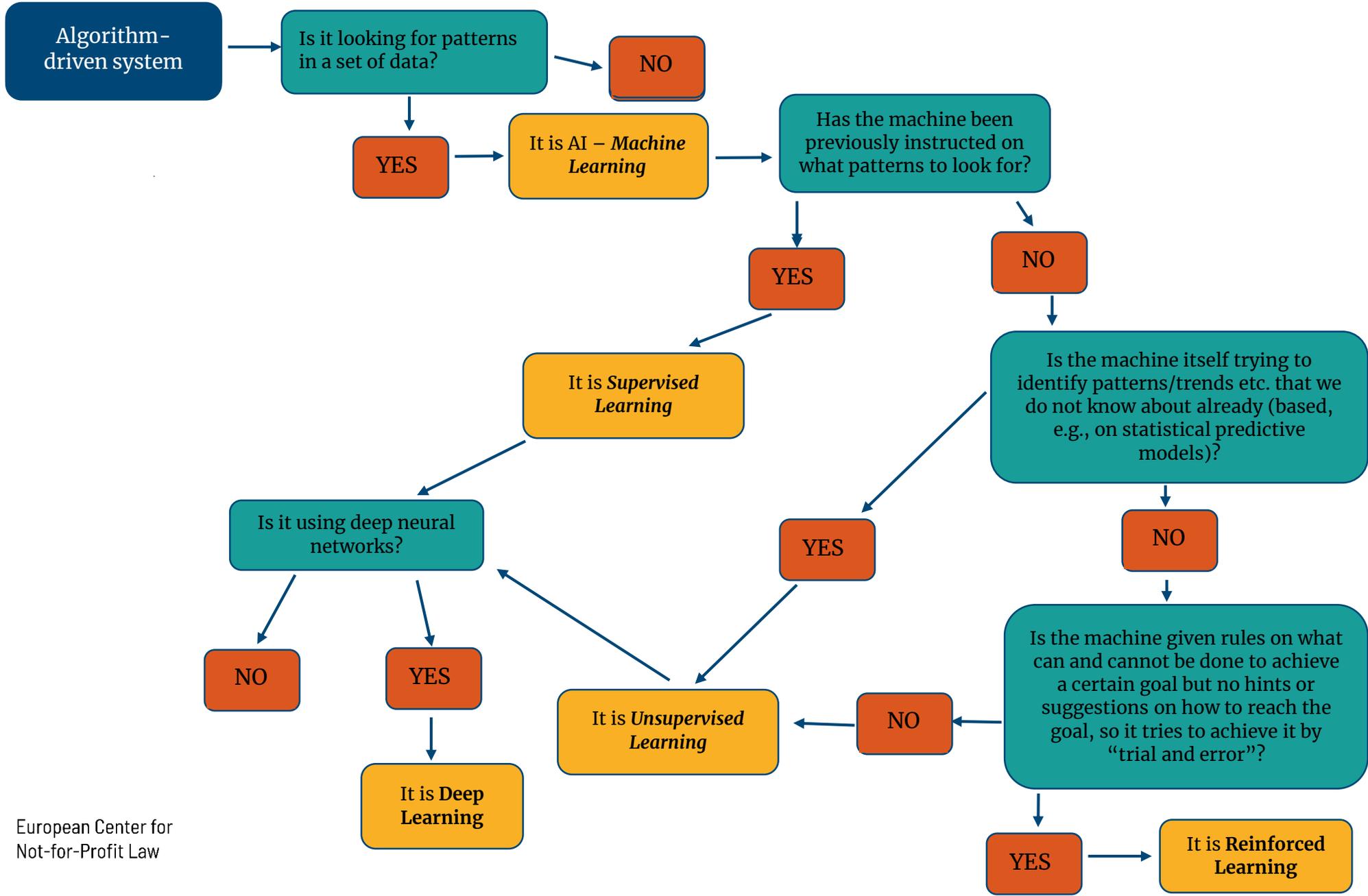
[@enablingNGOLaw](https://twitter.com/enablingNGOLaw)

AI: FACTS OR MYTHS

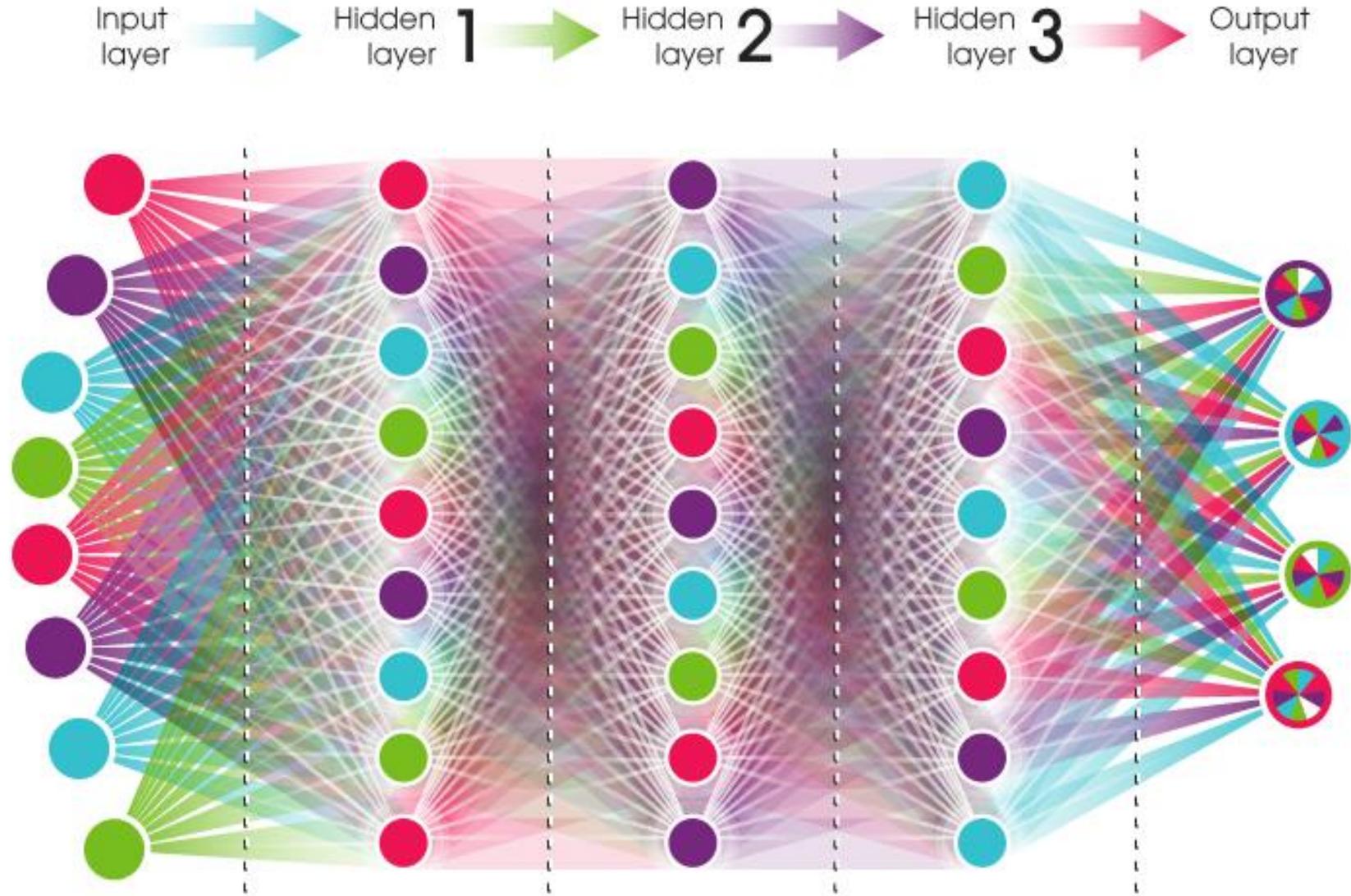
When we talk about AI, we refer to automated decision-making or decision-supporting systems operating via algorithms.







DEEP NEURAL NETWORK



AI: FACTS OR MYTHS

AI applications aim to simulate intelligent behaviour.



WHAT IS “INTELLIGENT BEHAVIOUR”?



Naturalist
Intelligence

Linguistic
Intelligence

Speculative
 (“Existential”)
Intelligence

Logical –
Mathematical
 (“Computational”)
Intelligence

Spatial
Intelligence

Bodily –
Kinesthetic
Intelligence

Musical
Intelligence

Interpersonal
 (“Empathetic”)
Intelligence



AI: FACTS OR MYTHS

AI applications can be for general purposes (“general AI”) or for specifically intended purposes (“narrow AI”). We already have both types of AI available nowadays.



General AI vs Narrow AI

- Systems with wide range of possible uses, both intended and unintended
 - (e.g., GPT-3 model generating extremist racist content)
- Applicable to different tasks in different fields without substantial modifications
- “Foundation models” = used as pre-trained models for more specialised AI systems (e.g., Natural Language Processing systems usable for question-answering, classification, translation, summarising, text prediction, speech recognition...)
- Eventually capable of learning and applying its intelligence to solve any problems

- Designed to perform specific tasks (e.g., facial recognition, image recognition system in healthcare, driving a car, selecting successful applicants, filtering spam, etc.)
- The only type of AI currently covered by the EU AI Act proposal



AI: FACTS OR MYTHS

AI applications always need data.
The more data available, the better.



AI systems training on datasets

Decision Process

Input
Data

E.g. your resume



Trained
System



Decision

Hire / no hire

Training Process

Training Data

E.g. dataset on historically
'successful' candidates

+

Algorithm

a computational process
or set of rules to be followed



Watch out for
"Less than one"-shot
learning model (aka,
"tiny synthetic
datasets")



AI: FACTS OR MYTHS

AI applications can be developed in such a way that they are never biased.



Two types of biases:

DATA BIAS

- Historical datasets (but times/context change)
- Patterns identified may even reveal “unconscious” prejudice

DESIGN BIAS

- Transference of developers' innate/unconscious prejudices in the coding of the algorithm

may reduce or amplify

Can
“synthetic
data”
solve this?



AI: FACTS OR MYTHS

AI systems can (and should) always be explainable.



AI Explainability

If the “logic” of the decisions for specific individuals cannot be explained



Contrary to the rule of law and human rights protection

“Black box”/lack of explainability of algorithm-based decisions



Contrary to “due process” and “reasoned judicial decisions”

This was at the core of the Siri case in The Netherlands (reason why the court struck down use of that AI-based application in the public sector).



AI DEFINITIONS

OECD – AI PRINCIPLES

An AI system is a **machine-based system** that can, for a **given set of human-defined objectives**, make **predictions, recommendations, or decisions influencing real or virtual environments**. AI systems are designed to operate with **varying levels of autonomy**.



AI DEFINITIONS

UNESCO Recommendation on Ethics of AI:

- Systems which have the **capacity to process data and information in a way that resembles intelligent behaviour**, and typically **includes aspects of reasoning, learning, perception, prediction, planning or control**.
- AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments.
- AI systems are **designed to operate with varying degrees of autonomy** by means of knowledge modelling and representation and by exploiting data and calculating correlations.
- AI systems **may include several methods, such as but not limited to**:(i) machine learning, including deep learning and reinforcement learning; (ii) machine reasoning, including planning, scheduling, knowledge representation and reasoning, search, and optimization.



AI DEFINITIONS

Council of Europe – Commissioner of Human Rights Recommendation (“Unboxing AI”):

- AI is used as an **umbrella term** to refer generally to a set of sciences, theories and techniques dedicated to **improving the ability of machines to do things requiring intelligence**.
- An AI system is a **machine-based system that makes recommendations, predictions or decisions for a given set of objectives**. It does so by:
 - (i) utilising machine and/or human-based inputs to perceive real and/or virtual environments;
 - (ii) abstracting such perceptions into models manually or automatically; and
 - (iii) deriving outcomes from these models, whether by human or automated means, in the form of recommendations, predictions or decisions.



AI DEFINITIONS

Draft EU AI Act:

Software that is developed with one or more of the techniques and approaches listed in Annex I and can, **for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments** they interact with.

Annex I:

- (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
- (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
- (c) Statistical approaches, Bayesian estimation, search and optimization methods.



Questions for break-out groups:

- **Can you provide examples of AI systems in your country (in the private and/or in public sector) based on the previous definitions and characteristics ?**
- **What benefits and/or challenges they present (if any) or can present in the short and longer term to individuals and society?**
- **Have you addressed these challenges yet in your NHRI (if yes, how/if not, how would you address them)?**



Sources/reading materials

- A People's Guide to AI (2018), <https://alliedmedia.org/wp-content/uploads/2020/09/peoples-guide-ai.pdf>
- Middlebury Institute of International Studies, *The Radicalization Risks of GPT-3 and Neural Language Models* (2020), https://www.middlebury.edu/institute/sites/www.middlebury.edu.institute/files/2020-09/The_Radicalization_Risks_of_GPT_3_and_Advanced_Neural_Language_Models_0.pdf?fv=Gx5PLI
- Future of Life Institute, *General Purpose AI and the AI Act*, <https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf>
- MIT Technology Review, Synthetic Data For AI (2022), <https://www.technologyreview.com/2022/02/23/1044965/ai-synthetic-data-2>
- The Economist, Briefing, *Huge “foundation models” are turbo-charging AI progress* (2022) <https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress>
- MIT Technology Review, *A radical new technique lets AI learn with practically no data* (2020), <https://www.technologyreview.com/2020/10/16/1010566/ai-machine-learning-with-tiny-data/>
- Netherlands, Court of The Hague, SyRI judgment (2020) <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878>
- AlgorithmWatch, *Automating Society Report 2020 - Mapping and analysis of automated decision-making (ADM) applications in the public policy sphere in Europe*, <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf>
- Council of Europe Commissioner For Human Rights Recommendation, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights* (2019), <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>
- OECD, Recommendation of the Council on Artificial Intelligence <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Proposal for a Regulation of the European Parliament and the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), (2021), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>

